



caBIG™

The cancer Biomedical Informatics Grid, or caBIG™, is a voluntary, virtual, biomedical informatics infrastructure that promises to speed progress in all aspects of cancer research including etiologic research, prevention, early detection and treatment. The National Cancer Institute (NCI) believes that caBIG™ will help redefine how cancer research is conducted, and eventually how patient care is provided.

This fact sheet provides an overview of the currently available caBIG™ tools and data sets and those pending release in 2005 and early 2006. Beyond these tools and data sets, caBIG™ is also delivering an expansive breadth of products such as OMB-derived Race and Ethnicity Standards, Candidate caBIG™ Data Standards, caBIG™ Compatibility Guidelines, project-specific white papers, end user materials for training and documentation, as well as information, educational materials, and template agreements relating to open source software licensing, publications, and other proprietary issues. All products, as they are completed, they will be available on the caBIG™ Web site at <http://caBIG.nci.nih.gov>.

caBIG™ Tools and Data Sets

Following are brief descriptions of each of the currently planned caBIG™ tools and data sets. Those available now are shown in blue.

Clinical Trials Solutions

Structured Protocol Representation

Standardized clinical trial protocol model

Adverse Events Reporting Tool

An open source software tool that is used to collect, process and report adverse events that occur during clinical trials

CTMS/CDUS Reporting Tool

A reporting interface that electronically translates data from various systems into required formats and submits to CDUS and CTMS

Trials Financial Billing Tool

Software that manages and monitors clinical trials expenditures and financial billing compliance

Trials Laboratory Interface

An interface that enables laboratory data across multiple systems and in multiple formats to be translated into common standards

C3D

Trials data capture application

C3PR

Clinical trial participant registry tool

Tissue Banks and Pathology Tools

caTISSUE Core

Database and specimen tracking system, part of the caTISSUE suite of modules

caTISSUE Annotation Module

Biological annotation and mapping system for specimens, part of the caTISSUE suite of modules

caTIES

Automated free-text pathology data extraction tool, part of the caTISSUE suite of modules

Integrative Cancer Research Tools

caWorkbench

Microarray analysis suite

Distance-Weighted Discrimination

Microarray data analysis integrator

Gene Pattern Bioinformatics Analysis Workflow

An analysis platform that supports multidisciplinary genomic research programs

Magellan

Tool for the analysis of heterogeneous data types (e.g., microarray)

Visual and Statistical Data Analyzer (VISDA)

Multivariate statistical visualization tool for the analysis of complex data



Cancer Molecular Pages System

A database and automated annotation system that combines automated computer-based annotations and automated data collection from experimental stations with a set of web-based visualization tools consuming data from caArray

FunctionExpress

Tool for integrated analysis and visualization of microarray data

Gene Ontology Miner (GOMiner)

Tool for aggregate analysis of gene sets

HapMap

caBIG™ accessible map of haplotypes in human genome

Promoter Database

A curated resource for the vertebrate transcription factor that binds sites and their corresponding regulatory regions

caArray

Cancer microarray data management system

Zebrafish Microarray Data Sharing

Microarray data from Zebrafish model system for human development

Pathways Tool Project

Pathway visualization tools

Quantitative Pathway Analysis in Cancer (QPACA)

Pathway modeling and analysis tool

Reactome (GKB) Data

Metabolic interactions database

Proteomics Laboratory Information Management System (LIMS)

A system used to track lab processes relevant to 2D-gel electrophoresis, with a schema to support addition of new data types as they emerge

Q5

Proteomics analysis tools

RProteomics

MALDI-TOF proteomics analysis tool

TrAPSS

Disease gene mutation discovery and analysis tool

Basic Architecture

caGrid Architecture

A prototype version of the basic Data Grid infrastructure for caBIG™, designed to satisfy the use case requirements from the various Domain Workspaces

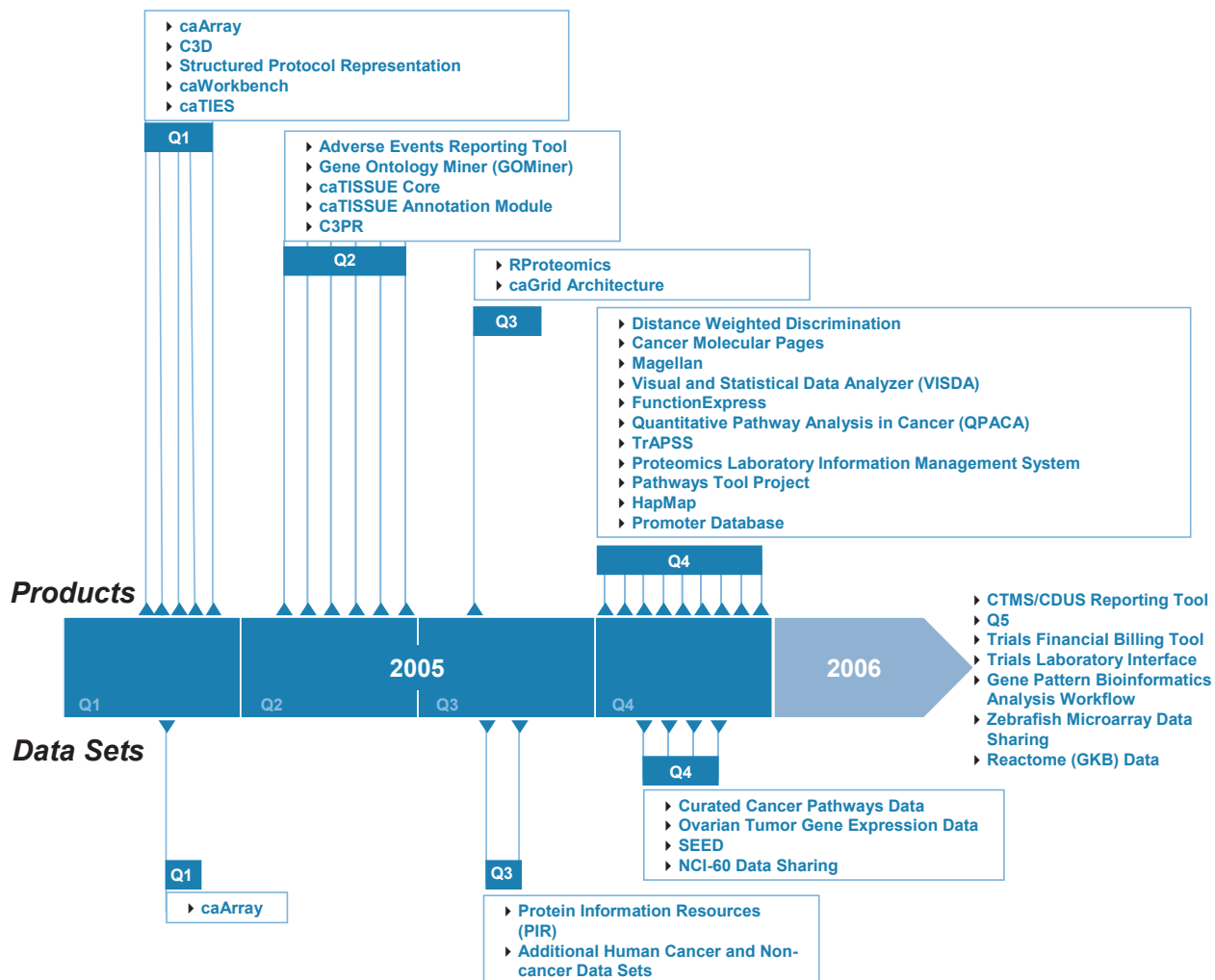
Data Sets

- Curated Cancer Pathways Data
- Ovarian Tumor Gene Expression Data
- SEED – Annotated human and cancer model organism genomes in a peer-to-peer annotation tool
- NCI-60 Data Sharing – Data sets generated from 60 human cancer cell lines
- Protein Information Resources (PIR) – Protein sequence and annotation data base
- Additional Human Cancer and Non-cancer Data Sets
- caARRAY – Cancer microarray data sets



Timeline

The diagram below provides a timeline representing the estimated delivery date for caBIG™ tools and data sets in the 2005 and early 2006 calendar years and highlights the rich inventory of new tools released each quarter. Not all tools to be released in 2006 have been identified; only tools that have been scheduled for release in early 2006 are shown. More detailed information on each of the tools can be found in the caBIG™ November Program Update at http://caBIG.nci.nih.gov/Program_Updates/caBIG_ProgramUpdate_November2004.pdf.



All project activities follow caBIG™'s guiding principles of open source, open access, open development, and federation; allowing the entire caBIG™ community and other cancer researchers to share applications, infrastructure, and data sets. Voluntary participation is welcome and encouraged, and will help ensure caBIG™'s long-term success. To find out more information or become involved with a Workspace or Working Group, please visit <http://caBIG.nci.nih.gov/workspaces>.

For more information about the caBIG™ initiative, please visit the caBIG™ Web site at <http://caBIG.nci.nih.gov>.

